Small Sample Properties of Alternative Tests for Martingale Difference Hypothesis

Amélie Charles, Olivier Darné, Jae Kim

Small Sample Properties of Alternative Tests for Martingale Difference Hypothesis

Amélie CHARLES

Audencia Nantes, School of Management

Olivier DARNÉ

LEMNA, University of Nantes

Jae H. KIM*

School of Economics and Finance, La Trobe University

August 30, 2010

*Corresponding author: School of Economics and Finance, La Trobe University, VIC 3086, Australia. Email: J.Kim@latrobe.edu.au.

Abstract

A Monte Carlo experiment is conducted to compare power properties of alternative tests for the martingale difference hypothesis. Overall, we find that the wild bootstrap automatic variance ratio test shows the highest power against linear dependence; while the generalized spectral test performs most desirably under nonlinear dependence.

Keywords: Martingale difference hypothesis; variance ratio test; portmanteau test; spectral test;

JEL Classification: C12; C14.

1 Introduction

Testing for the martingale difference hypothesis (MDH) is central in many economic and finance studies, such as market efficiency, rational expectations, and optimal consumption smoothing. A martingale difference sequence (MDS) has no dependence in mean, conditional on its own history, implying that it is purely non-predictable from its own past. Escanciano and Lobato (2009b) provides an informative review of the statistical tests for the MDH. The tests widely used in empirical applications include the portmanteau test (Ljung and Box, 1978) and variance ratio test (Lo and MacKinlay, 1988), which are based on the linear measures of dependence. Notable recent contributions to this category of the MDH tests include the automatic portmanteau (AQ) test of Escanciano and Lobato (2009a); and automatic variance ratio (AVR) test of Kim (2009) extending the earlier work of Choi (1999). The other category of the MDH tests adopts nonlinear measures of dependence, which include the generalized spectral (GS) test of Escanciano and Velasco (2006) and the consistent tests of Dominguez and Lobato (2003; DL tests).

In this paper, we conduct an extensive Monte Carlo study to compare small sample properties of these alternative tests. We aim to provide a guideline as to which tests should be preferred in practical applications. We compare their power properties under a wide range of linear and nonlinear models. The next section provides a brief review of these tests, and Section 3 reports the Monte Carlo results.

2 Tests for Martingale Difference Hypothesis

To conserve space, only brief details of the tests are provided. Let $\{Y_t\}_{-\infty}^{\infty}$ denote a real-valued stationary time series. Under the MDH, $E[Y_t|I_{t-1}] = \mu$; or equivalently $E[(Y_t - \mu)\omega(I_{t-1})] = 0$, where $I_t = \{Y_t, Y_{t-1}, ...\}$ is the information set at time t and $\omega(\cdot)$ is a weighting function. Here, $\omega(I_{t-1})$ represents any linear or nonlinear transformation of the past. Depending on the choice of this weighting function, the tests are classified into those based on linear or nonlinear measures of dependence.

2.1 Tests based on linear measures of dependence

When the weighting function takes the linear form, i.e., $\omega(I_{t-1}) = Y_{t-i}$ for some $i \geq 1$, the MDH implies $\gamma_i \equiv E[(Y_t - \mu)(Y_{t-i} - \mu)] = 0$. The most popular tests are the portmanteau and variance ratio tests for $H_0 : \rho(i) \equiv \gamma_i/\gamma_0 = 0$. The original portmanteau test statistic is written as

$$Q_p = T \sum_{i=1}^{p} \widehat{\rho}^2(i), \tag{1}$$

where $\hat{\rho}(i)$ is the sample estimate of $\rho(i)$ and T is the sample size. When Y_t has conditional heteroscedasticity, Lobato et al. (2001) propose the robustified statistic of the form

$$Q_p^* = T \sum_{i=1}^{p} \widetilde{\rho}^2(i), \tag{2}$$

where $\widetilde{\rho}(i) = \widehat{\gamma}^2(i)/\widehat{\tau}(i)$, $\widehat{\gamma}(i)$ is the sample autocovariance of Y_t, and $\widehat{\tau}(i)$ is the sample autocovariance of Y_t^2.

To avoid an *ad hoc* selection of p, Escanciano and Lobato (2009a) propose an automatic test where the optimal value of p is determined by a fully data-dependent procedure. The test statistic, which asymptotically follows the χ_1^2 distribution, is written as

$$AQ = Q^*_{\widetilde{p}} \tag{3}$$

where $\widetilde{p} = \min\{p : 1 \leq p \leq d; L_p \geq L_h, h = 1, 2, ..., d\}$ and d is a fixed upper bound, while $L_p = Q_p^* - \pi(p, T)$, where the penalty term $\pi(p, T, q) = p \log(T)$ if $\max_{1 \leq i \leq d} \sqrt{T} |\widetilde{\rho}(i)| \leq \sqrt{2.4 \log(T)}$ and $\pi(p, T, q) = 2p$ if otherwise. Note that the penalty term is a balance between AIC and BIC.

The variance ratio test can be written as

$$\widehat{VR}(k) = 1 + 2 \sum_{i=1}^{k-1} (1 - \frac{i}{k}) \widehat{\rho}(i), \tag{4}$$

where k denotes the holding period. Choi (1999) proposes an automatic variance ratio (AVR) test where k is chosen optimally using a fully data-dependent method based on Andrews (1991). Kim (2009) finds that small sample properties of Choi's (1999) test can be substantially improved, by employing the wild bootstrap. Let the AVR test statistic with the optimal choice of k be denoted as $AVR(k^*)$. Kim's (2009) wild bootstrap AVR test is conducted in three stages as follows:

1. Form a bootstrap sample of size T as $Y_t^* = \eta_t Y_t$ $(t = 1, ..., T)$, where η_t is random variable with zero mean and unit variance;
2. Calculate $AVR^*(k^*)$, the $AVR(k^*)$ statistic calculated from $\{Y_t^*\}_{t=1}^T$;
3. Repeat 1 and 2 B times, to produce the bootstrap distribution of the AVR statistic $\{AVR^*(k^*; j)\}_{j=1}^B$.

To test for H_0 against the two-tailed alternative, the bootstrap p-value is calculated as the proportion of the absolute values of $\{AVR^*(k^*; j)\}_{j=1}^B$ greater than the absolute value of the observed statistic $AVR(k^*)$. For η_t, we use the two point distribution given in Escanciano and Velasco (2006; p.164).

Note that the AQ and AVR tests can be inconsistent against nonlinear alternatives. The latter include a time series which are serially uncorrelated but dependent. The AVR test also has a serious theoretical limitation of being inconsistent even for some linear models, which occurs when the autocorrelations of different signs cancel out (see Escanciano and Lobato; 2009b, p.979). The VR test statistic in (4) is not robust to heteroskedasticity, although its wild bootstrap version provides statistical inference robust to heteroskedasticity (see, Kim; 2006).

2.2 Tests based on nonlinear measures of dependence

For the case of general nonlinear weighting function, popular choices have been exponential function and indicator function. The former is to detect the general nonlinear conditional mean dependence, and the latter to test for no directional predictability.

Escanciano and Velasco (2006) express the null of the MDH in a form of pairwise regression function. That is, $H_0 : m_j(y) = 0$, where $m_j(y) = E(Y_t - \mu | Y_{t-j} = y)$, against $H_1 : P[m_j(y) \neq 0] > 0$ for some j. They note that the above null hypothesis is consistent with the exponential weighting function such that

$$\gamma_j(x) \equiv E[(Y_t - \mu)e^{ixY_{t-j}}] = 0,$$

where $\gamma_j(x)$ represents an autocovariance measure in a nonlinear framework with x being any real number. Escanciano and Velasco (2006) propose the use of the generalized spectral distribution function, whose sample estimate is written as

$$\widehat{H}(\lambda, x) = \widehat{\gamma}_0(x)\lambda + 2\sum_{j=1}^{\infty}(1 - \frac{j}{T})\widehat{\gamma}_j(x)\frac{\sin(j\pi\lambda)}{j\pi},$$

where $\widehat{\gamma}_j(x) = (T-j)^{-1}\sum_{t=1+j}^{T}(Y_t - \overline{Y}_{T-j})e^{ixY_{t-j}}$ and $\overline{Y}_{T-j} = (T-j)^{-1}\sum_{t=1+j}^{T} Y_t$. Under the null hypothesis, $\widehat{H}(\lambda, x) = \widehat{\gamma}_0(x)\lambda \equiv \widehat{H}_0(\lambda, x)$, and the test statistic for H_0 is written as

$$\begin{aligned} S_T(\lambda, x) &= (0.5T)^{1/2}\{\widehat{H}(\lambda, x) - \widehat{H}_0(\lambda, x)\} \\ &= \sum_{j=1}^{T-1}(T-j)^{0.5}\widehat{\gamma}_j(x)\frac{\sqrt{2}\sin(j\pi x)}{j\pi}. \end{aligned}$$

To evaluate the value of S_T for all possible values of λ and x, Escanciano and Velasco (2006) use the Cramer-von Mises norm to obtain the statistic

$$D_T^2 = \sum_{j=1}^{T-1}\frac{(T-j)}{(j\pi)^2}\sum_{t=j+1}^{T}\sum_{s=j+1}^{T}\exp(-0.5(Y_{t-j} - Y_{s-j})^2). \tag{5}$$

Dominguez and Lobato (2003) consider the case of indicator weighting function and propose the (DL) tests based on Cramer-von Mises (CvM) and Kolmogorov-Smirnov (KS) statistics, which can be written as

$$CvM_{T,p} = \frac{1}{\widehat{\sigma}^2T^2}\sum_{j=1}^{T}[\sum_{t=1}^{T}(Y_t - \overline{Y})1(\widetilde{Y}_{t,p} \leq \widetilde{Y}_{j,p})]^2; \tag{6}$$

$$KS_{T,p} = \max_{1\leq i\leq T}|\frac{1}{\widehat{\sigma}\sqrt{T}}\sum_{t=1}^{T}(Y_t - \overline{Y})1(\widetilde{Y}_{t,p} \leq \widetilde{Y}_{j,p})|, \tag{7}$$

where $\widetilde{Y}_{t,p} = (Y_{t-1}, ..., Y_{t-p})$ and p is a positive integer.

The GS and DL test statistics given in (5) to (7) do not possess the standard asymptotic distributions. To implement the tests in finite samples, the above authors recommend the use of the wild bootstrap. That is, the p-value of the test can be obtained from the wild bootstrap distribution, as described in Section 2.1 for the AVR test. The DL tests are conditional on finite-dimensional information set, requiring the choice of lag order p; while the GS exploits infinite-dimensional information set. As noted in Escanciano and Velasco (2006), the GS test is only pairwise consistent, but is inconsistent against pairwise MDS which are non-MDS.

3 Monte Carlo Simulations

We only report power properties (the probability of rejection under H_1), because we find no evidence of size distortion for all tests, except for the AQ test **which is** slightly oversized only when the sample size is as small as 100. We consider a number of linear and nonlinear models. For the former,

- AR(1) model: $Y_t = 0.1Y_{t-1} + Z_t$, and $Y_t = 0.1Y_{t-1} + V_t$
- ARFIMA model: $(1 - L)^{0.1}Y_t = Z_t$; and $(1 - L)^{0.1}Y_t = V_t$;
- **The sum of a white noise and the first difference of a stationary autoregressive process of order one (NDAR)**: $Y_t = \epsilon_t + X_t - X_{t-1}$ with $X_t = 0.85X_{t-1} + u_t$,

where $Z_t = \epsilon_t\sigma_t$ with $\sigma_t^2 = 0.001 + 0.90\sigma_{t-1}^2 + 0.09\epsilon_{t-1}^2$ (**i.e. GARCH(1,1) errors**); $V_t = \exp(0.5h_t)\epsilon_t$ with $h_t = 0.95h_{t-1} + u_t$ (**i.e. stochastic volatility (SV) errors**); ϵ_t and u_t are independent i.i.d. N(0,1). For nonlinear models, we consider four models used by Escanciano and Velasco (2006), which include

- Bilinear model: $Y_t = \epsilon_t + 0.25\epsilon_{t-1}Y_{t-1} + 0.15\epsilon_{t-1}Y_{t-2}$;
- TAR(1) model: $Y_t = -0.5Y_{t-1} + \epsilon_t$ if $Y_t \geq 1$ and $Y_t = 0.4Y_{t-1} + \epsilon_t$ if $Y_t < 1$;
- **Non-linear moving average model (NLMA)**: $Y_t = \epsilon_{t-1}\epsilon_{t-2}(\epsilon_{t-2} + \epsilon_t + 1)$; and
- **First order exponential autoregressive model (EXP(1))**: $Y_t = 0.6Y_{t-1}\exp(-0.5Y_{t-1}^2) + \epsilon_t$.

The sample size considered are 100, 300, and 500. The number of bootstrap iterations (B) for the AVR, DL and GS tests are set to 500, and the number of Monte Carlo trials to 1000. For the DL tests, we only report the case where $p = 1$, since they show lower power when $p > 1$.

Table 1 reports the power of the linear tests under linear models. The AVR and AQ tests perform similarly, but the former shows higher power in most cases. Table 2 reports the power of the nonlinear tests under linear models. Under GARCH errors, the GS test shows higher power than the DL tests, but the DL tests are more powerful under the SV errors. Table 3 reports the power of the linear tests under nonlinear models. The AVR test performs much better than the AQ test, showing high power especially for the bilinear, EXP(1) and TAR(1) models. Table 4 reports the power of the nonlinear tests under nonlinear models. The GS test is more powerful than the DL for the bilinear and TAR(1) models. For NLMA and EXP(1), both perform similarly, but the DL tests tend to show higher power. As expected, the nonlinear tests are more powerful than the linear tests under nonlinear models; but the reverse tends to be the case under linear models.

Overall, it is found that the AVR and GS tests show excellent power against a wide range of linear and nonlinear models, with no size distortion. The DL tests also show satisfactory performance, being more powerful than the GS test under SV errors. Although the AVR test is not designed to detect nonlinear dependency, it shows good power properties against a range of nonlinear models. Since it is often uncertain in practice whether the nature of dependency is linear or nonlinear, the use of the AVR, along with the DL and GS tests, is strongly recommended. As a further note, we find that wild bootstrapping does not improve the power of the AQ test (the details are not reported).

References

[1] Andrews, D.W.K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica, 58, 817–858.

[2] Choi, I. 1999. Testing the random walk hypothesis for real exchange rates. Journal of Applied Econometrics 14, 293-308.

[3] Dominguez, M. A., Lobato, I.N. 2003. Testing the Martingale Difference Hypothesis. Econometrics Review 22, 371-377.

[4] Escanciano, J.C., Lobato, I.N. 2009a. An automatic portmanteau test for serial correlation. Journal of Econometrics 151, 140–149.

[5] Escanciano, J.C., Lobato, I.N. 2009b. Testing the Martingale Hypothesis. In Patterson K. and Mills T.C. (eds) Palgrave Handbook of Econometrics, Palgrave, MacMillan.

[6] Escanciano, J.C., Velasco, C. 2006. Generalized spectral tests for the martingale difference hypothesis. Journal of Econometrics 134, 151–185.

[7] Kim, J.H. 2006. Wild bootstrapping variance ratio tests. Economics Letters 92, 38-43.

[8] Kim, J.H. 2009. Automatic variance ratio test under conditional heteroskedasticty. Finance Research Letters 3, 179-185.

[9] Ljung, G.M., Box, G.E.P. 1978. On a measure of lack of fit in time series models. Biometrika 65, 297–303.

[10] Lo, A.W., MacKinlay, A.C. 1988. Stock market prices do not follow random walk: Evidence from a simple specification test. The Review of Financial Studies 1, 41-66.

[11] Lobato, I.N., Nankervis, J.C., Savin, N.E. 2001. Testing for autocorrelation using a modified Box-Pierce Q test. International Economic Review 42, 187–205.